# Join Graph Decomposition Bounds for Influence Diagrams

UCIRVINE

Junkyu Lee, Alexander Ihler, Rina Dechter

University of California, Irvine {junkyul, ihler, dechter}@ics.uci.edu

## Summary

**A new decomposition method for bounding the MEU**

- **Join graph decomposition bounds for IDs (JGDID)**
  - Approximate inference algorithm for influence diagram
  - Proposed method is based on the valuation algebra
  - Exploits local structure of influence diagrams
  - Extends dual decomposition for MMAP

**Significant improvement in upper bounds compared with earlier works**

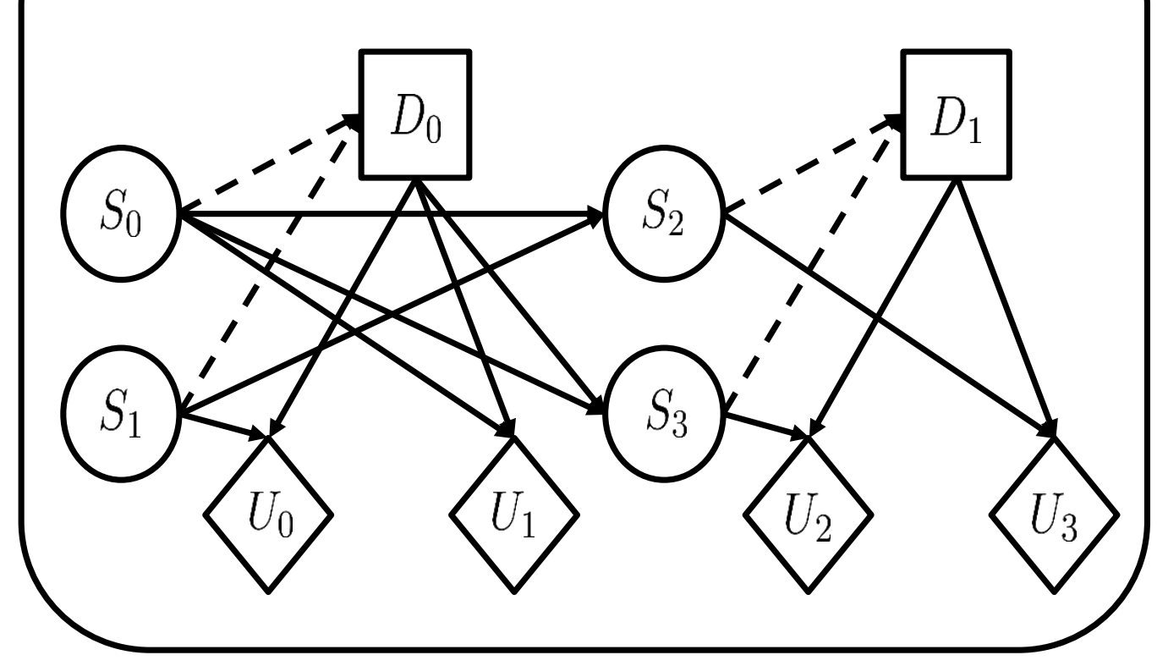
- **Translation based methods**
  - Pure/interleaved MMAP translation + MMAP inference
- **Direct relaxation methods**
  - mini-bucket scheme with valuation algebra
  - relaxing non-anticipativity constraint

## Background – Influence Diagram

[howard and Matheson, 2005]

❖ **A graphical model for sequential decision-making under uncertainty with perfect recall**

**Factored MDP as an ID**



**Chance variables** $\mathbf{C} = \{S_0, S_1, S_2, S_3\}$

**Decision variables** $\mathbf{D} = \{D_0, D_1\}$

**Probability functions** $\mathbf{P} = \{P(S_0), \cdots, P(S_3|pa(S_3))\}$

**Utility functions** $\mathbf{U} = \{U_0(pa(U_0), \cdots, U_3(pa(U_3))\}$

**Partial ordering constraint** $\mathcal{O} = \{S_0, S_1\} \prec \{D_0\} \prec \{S_2, S_3\} \prec \{D_1\}$

**Policy functions** $\Delta_0 := R(D_0) \mapsto D_0 \quad \Delta_1 := R(D_1) \mapsto D_1$

**Task – compute MEU and optimal policy**

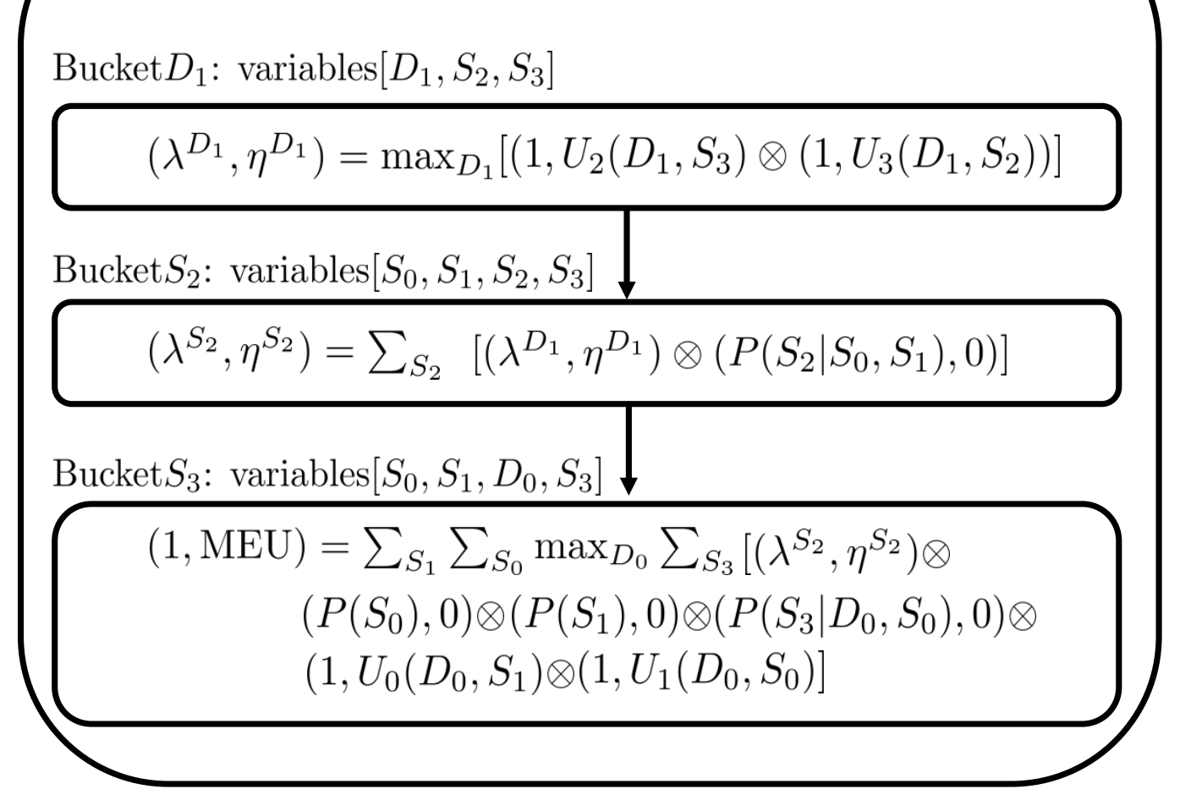
$$\textstyle\sum_{S_0,S_1} \max_{D_0} \sum_{S_2,S_3} \max_{D_1} [\prod_{P_i \in \mathbf{P}} P_i][\sum_{U_i \in \mathbf{U} U_i}][\prod_{\Delta_i \in \mathbf{\Delta}} \Delta_i]$$

## Background – Valuation Algebra

[Jensen 1994, Maua 2012]

❖ **Algebraic framework for computing expected utility value (a.k.a. potential)**

**Variable elimination with VA**



$\text{Bucket}D_1\text{: variables}[D_1, S_2, S_3]$

$(\lambda^{D_1}, \eta^{D_1}) = \max_{D_1}[(1, U_2(D_1, S_3) \otimes (1, U_3(D_1, S_2))]$

$\text{Bucket}S_2\text{: variables}[S_0, S_1, S_2, S_3]$

$(\lambda^{S_2}, \eta^{S_2}) = \sum_{S_2} [(\lambda^{D_1}, \eta^{D_1}) \otimes (P(S_2|S_0, S_1), 0)]$

$\text{Bucket}S_3\text{: variables}[S_0, S_1, D_0, S_3]$

$(1, \text{MEU}) = \sum_{S_1} \sum_{S_0} \max_{D_0} \sum_{S_3} [(\lambda^{S_2}, \eta^{S_2}) \otimes (P(S_0), 0) \otimes (P(S_1), 0) \otimes (P(S_3|D_0, S_0), 0) \otimes (1, U_0(D_0, S_1) \otimes (1, U_1(D_0, S_0)]$

**Valuation:** $\Psi(\mathbf{X}) := (P(\mathbf{X}), V(\mathbf{X}))$

$P(\mathbf{X})$ **probability** $V(\mathbf{X})$ **expected utility**

**Combination**

$\Psi_1 := (P_1, V_1) \quad \Psi_2 := (P_2, V_2)$

$\Psi_1 \otimes \Psi_2 := (P_1 P_2, P_1 V_2 + P_2 V_1)$

**Marginalization**

$\Psi := (P(\mathbf{X}), V(\mathbf{X})) \quad \mathbf{Y} \subseteq \mathbf{X}$

$\sum_{\mathbf{Y}} \Psi := (\sum_{\mathbf{Y}} P, \sum_{\mathbf{Y}} V)$

$\max_{\mathbf{Y}} \Psi := (\max_{\mathbf{Y}} P, \max_{\mathbf{Y}} V)$

**MEU Query in VA**

$\mathbf{\Psi} := \{(P_i, 0) | P_i \in \mathbf{P}\} \cup \{(1, U_i) | U_i \in \mathbf{U}\}$

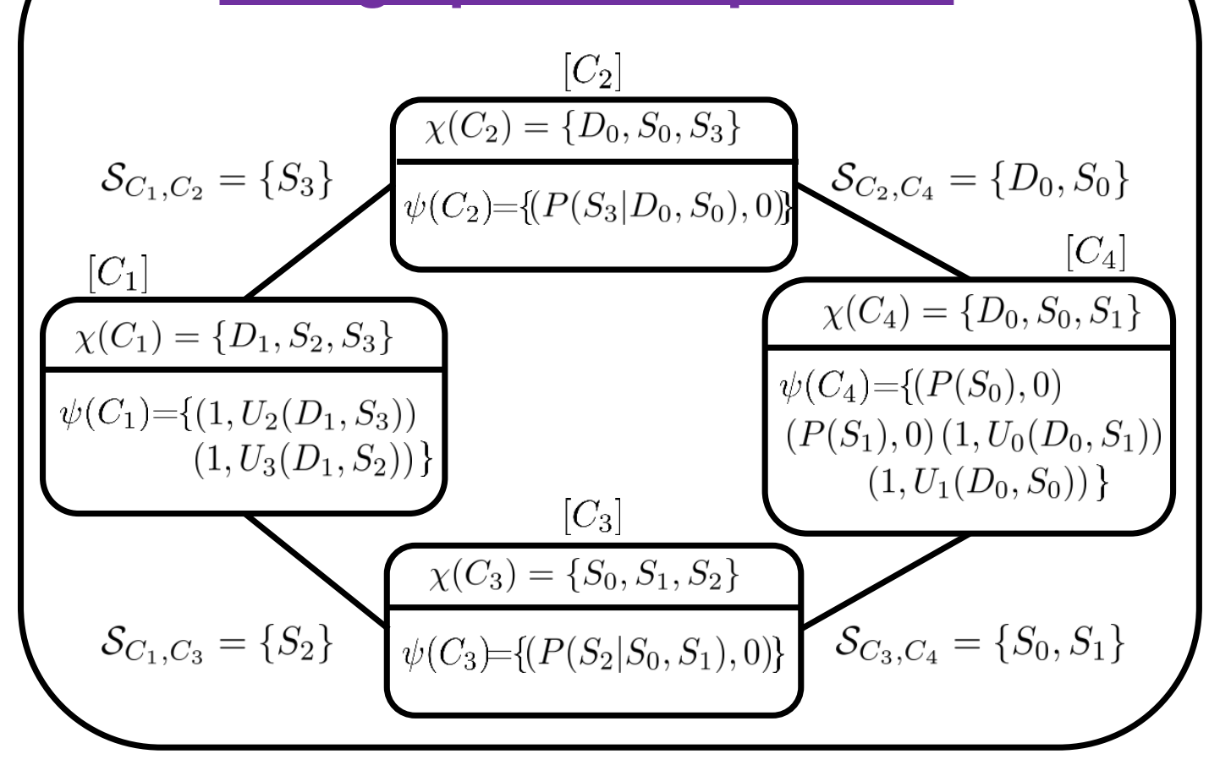
$\sum_{S_0,S_1} \max_{D_0} \sum_{S_2,S_3} \max_{D_1} \otimes_{i \in \mathcal{I}_\Psi} \Psi_i$

## Background – Join Graph Decomposition

[Mateescu, Kask, Gogate, Dechter 2009]

❖ **Approximation scheme that decomposes a Join tree by limiting the maximum cluster size**

**Join graph decomposition**



**Join Graph Decomposition of Influence Diagrams**

Join Graph $\mathcal{G}_J := (\mathcal{C}, \mathcal{S})$ with set of nodes $\mathcal{C}$ and edges $\mathcal{S}$

Node labeling function $\chi(C)$ maps each node to a subset of variables $\mathbf{X}_C$ and **assign valuations exclusively to a node**

Separator $\mathcal{S}_{C_i,C_j}$ intersection of the variables between $C_i$ and $C_j$

Join graph decomposition satisfies **running intersection property**

## Decomposition Bounds for IDs

❖ **(Definition) Powered-sum elimination for a valuation algebra**

- **generalize elimination operator by Lᵖ-norm**

Given $\Psi(\mathbf{X}) := (P(\mathbf{X}), V(\mathbf{X}))$ over $\mathbf{X}$

$$\textstyle\sum_X^{(w,A)} \Psi(X) := (\sum_X^w P(X), \sum_X^w h(P(X), V(X), A)) \otimes (1, -A)$$

with $\sum_X^w f(X) = [\sum_X |f(X)|^{1/w}]^w \quad 0 \le w \le 1$

$$h(P(X), V(X), A) = \begin{cases} P(X)(\frac{V(X)}{P(X)} + A) & \text{if } \frac{V(X)}{P(X)} + A > 0 \\ 0, & \text{otherwise.} \end{cases}$$

❖ **(Theorem) Decomposition Bounds for IDs**

- **decomposition bounds interchange elimination and combination**

Given an ID $\mathcal{M}' := \langle \mathbf{X}, \mathbf{\Psi}, \mathcal{O} \rangle$, the MEU can be bounded by

$$\sum_{\mathcal{O}}^{\mathbf{w}} \otimes_{\alpha \in \mathcal{I}_{\mathbf{\Psi}}} \Psi_\alpha(\mathbf{X}_\alpha) \le \otimes_{\alpha \in \mathcal{I}_{\mathbf{\Psi}}} \sum_{\mathcal{O}}^{(\mathbf{w}^\alpha, \mathbf{A})} \Psi_\alpha(\mathbf{X}_\alpha).$$

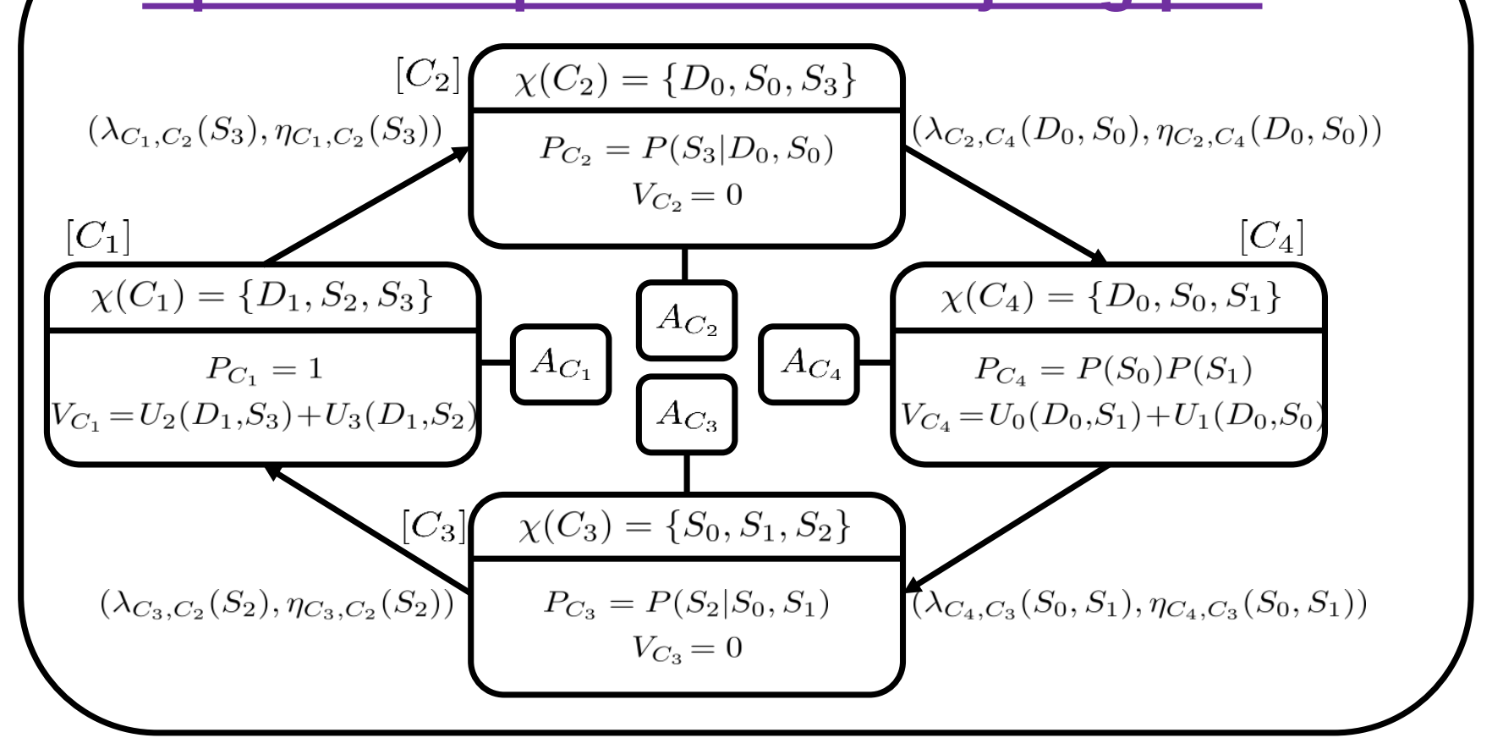
with $w_i = \sum_{\alpha \in \mathcal{I}_\Psi} w_i^\alpha$ for $\mathbf{w}$ and $\mathbf{w}^\alpha \ \forall \alpha \in \mathcal{I}_{\mathbf{\Psi}}$

$\mathbf{A} := \{A_\alpha | \forall \alpha \in \mathcal{I}_{\mathbf{\Psi}}\}$

## Parameterized Bounds over Join Graph

❖ **(Proposition) Parameterized decomposition bounds by introducing optimization parameters relative to a join graph**

**Optimization parameters over join grpah**



**Utility constant parameters over nodes** $A_{C_i}$

**Cost shifting valuations over edges** $\delta_{C_i,C_j} = (\lambda_{C_i,C_j}, \eta_{C_i,C_j})$

**Weights from Lᵖ norm over the variables** $w_X = \sum_{C_i \in \mathcal{C}} w_X^{C_i}$

**Parameterized decomposition bounds**

**probability**

$$L_{\text{MMAP}} = \textstyle\prod_{C_i \in \mathcal{C}} \sum_{\mathcal{O}}^{\mathbf{w}^{C_i}} P'_{C_i}$$

**expected utility**

$$L_{\text{MEU}} = \sum_{C_i \in \mathcal{C}} [\sum_{\mathcal{O}}^{\mathbf{w}^{C_i}} h_{C_i}(P'_{C_i}, V'_{C_i})][\prod_{C_j \ne C_i} \sum_{\mathcal{O}}^{\mathbf{w}^{C_j}} P'_{C_j}] - A_{C_i}$$

with re-parameterized functions

$$P'_{C_i} = P_{C_i} \textstyle\prod_{(C_i,C_j) \in \mathcal{S}} \lambda_{C_i,C_j}$$

$$V'_{C_i} = P'_{C_i} [\frac{V_{C_i}}{P_{C_i}} + \sum_{(C_i,C_j) \in \mathcal{S}} \frac{\eta_{C_i,C_j}}{\lambda_{C_i,C_j}}]$$

## References & Acknowledgement

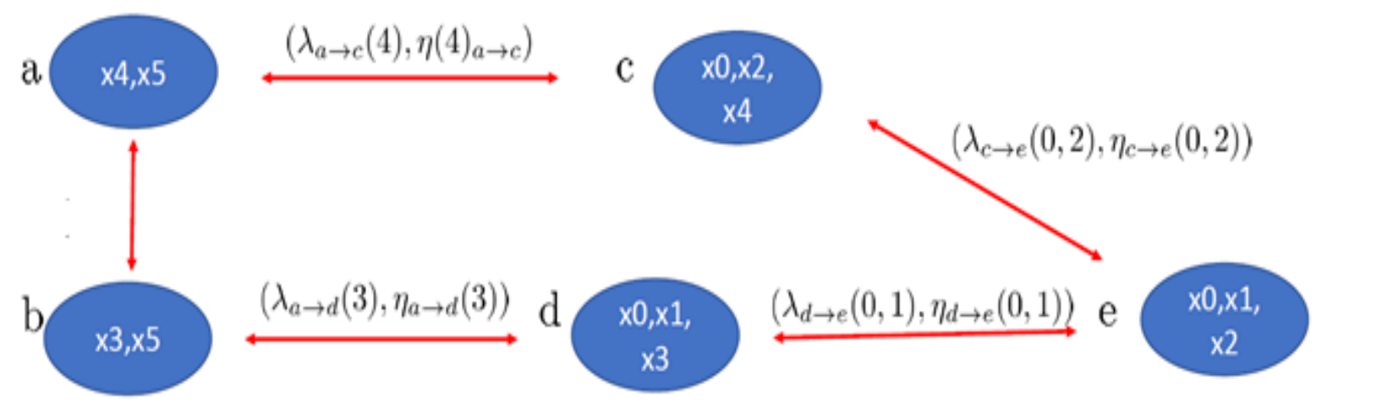
- [Dechter 1999] Bucket elimination: A unifying framework for reasoning. Artificial Intelligence.
- [Dechter 2000] An anytime approximation for optimizing policies under uncertainty. AIPS-2000.
- [Dechter and Rish 2003] Mini-buckets: A general scheme for bounded inference. Journal of the ACM.
- [Howard and Matheson 2005] Influence diagrams. Decision Analysis.
- [Jensen, Jensen, and Dittmer 1994] From influence diagrams to junction trees. UAI-1994.
- [Kivinen and Warmuth 1997] Exponentiated gradient versus gradient descent for linear predictors. Information and Computation.
- [Liu and Ihler 2011] Bounding the partition function using Hölder's inequality. ICML 2011.
- [Liu and Ihler 2012] Belief propagation for structured decision making. UAI-2012.
- [Marinescu, Dechter, and Ihler 2014] AND/OR search for marginal MAP. UAI-2014.
- [Mateescu, Kask, Gogate, and Dechter 2010] Join-graph propagation algorithms. JAIR.
- [Mauá 2016] Equivalences between maximum a posteriori inference in Bayesian networks and maximum expected utility
- [Mauá and Zaffalon 2012] Solving limited memory influence diagrams. Journal of Artificial Intelligence Research.
- [Nilsson and Hohle 2001] Computing bounds on expected utilties for optimal policies based on limited information. Technical Report.
- [Ping, Liu, and Ihler 2015] Decomposition bounds for marginal MAP. NIPS-2015.
- [Sontag Globerson, and Jaakkola 2011] Introduction to dual decomposition for inference. Optimization for Machine Learning.
- [Wright and Nocedal 1999] Numerical optimization. Springer Science.
- [Yuan, and Hansen 2010] Solving multistage influence diagrams using branch-and-bound search. UAI-2010.

### Acknowledgement

This work was supported in part by NSF grants IIS-1526842 and IIS-1254071, the US Air Force (Contract FA9453-16-C-0508) and DARPA (Contract W911NF-18-C-0015).

## Message Passing Algorithm (JGDID)

```
Algorithm 1 Join Graph Decomposition for IDs (JGDID)
Require: Influence diagram $\mathcal{M}' = \langle \mathbf{X}, \mathbf{\Psi}, \mathcal{O} \rangle$, initial weights
    $w_i$ associated with a variable $X_i \in \mathbf{X}$, $i$-bound, total iteration limit $M_1$,
    iteration limit $M_2$ for updating weights and costs before updating utility constants.
Ensure: an upper bound of the MEU, $L_{\text{MEU}}$,
generate a join graph decomposition $\mathcal{D} = \langle \mathcal{G}_J(\mathcal{C}, \mathcal{S}), \chi, \psi \rangle$
    by MBE with $i$-bound and assign valuations to nodes by labeling function $\psi$.
execute single pass cost-shifting by messages generated by
    MBE algorithm based on the valuation algebra (MBE-VA)
initialize weights $\mathbf{w}^{C_i}, \forall C_i \in \mathcal{C}$ by uniform weights.
$iter$=0, $L_{\text{MEU}} = \inf$
while $iter < M_1$ or $L_{\text{MEU}}$ is not converged do
  for each variable $X_i \in \mathbf{X}$ do
    $L_{\text{MEU}} \leftarrow \min(L_{\text{MEU}}, \text{UPDATE-WEIGHTS}(\mathcal{G}_J, X_i))$
  end for
  for each edge $(C_i, C_j) \in \mathcal{S}$ do
    $L_{\text{MEU}} \leftarrow \min(L_{\text{MEU}}, \text{UPDATE-COSTS}(\mathcal{G}_J, \{C_i, C_j\}))$
  end for
  if $iter > M_2$ then
    for each node $C_i \in \mathcal{C}$ do
      $L_{\text{MEU}} \leftarrow \min(L_{\text{MEU}}, \text{UPDATE-UTIL-CONST}(\mathcal{G}_J, C_i))$
    end for
  end if
  $iter = iter + 1$
end while
```



**Initialization**

- **Join graph decomposition** [Mateescu, et la 2010]
- **Mini-bucket elimination** [Dechter and Rish, 2003]

**Outer optimization by Block coordinate method**

- **optimize subset of parameters for nonconvex** $L_{\text{MEU}}$

**Inner optimization by first order methods**

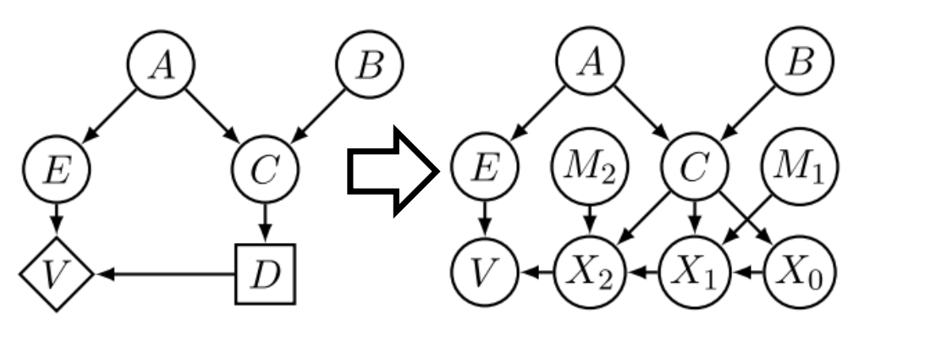
- **Weights per variable:** exponentiated gradient descent [Kivinen and Warmuth, 1997]
- **Cost per edges:** gradient descent
- **Utility constants per nodes:** gradient descent

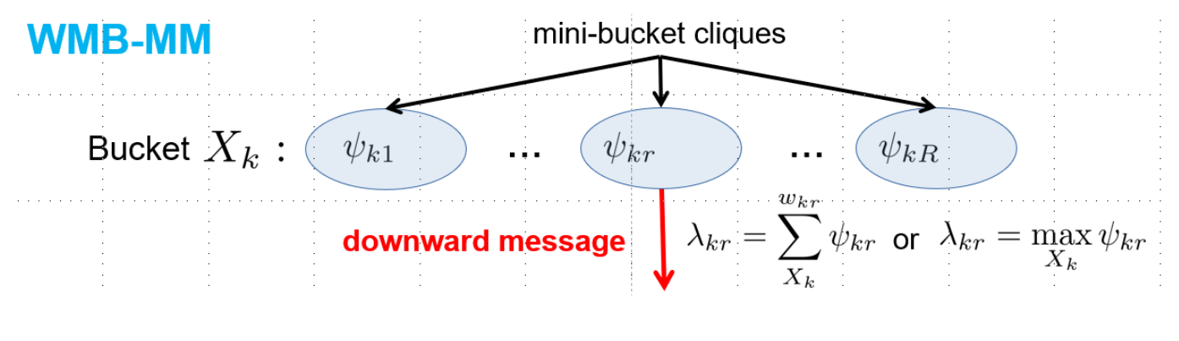
$$\mathbf{x}^{t+1} = \mathbf{x}^t - s \cdot [\nabla f(\mathbf{x}^t)]$$

## Earlier Works

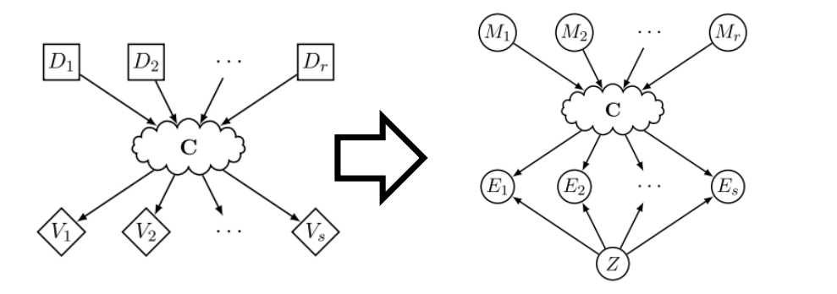
❖ **MMAP translation + approximate MMAP inference**

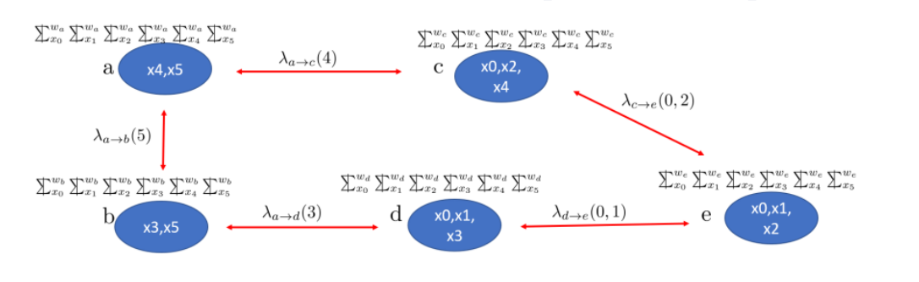
**Reduction of ID to MMAP** [Maua 2016] + **WMBMM (Weighted Mini-Bucket with Moment Matching)** [Liu, Ihler 2011, Marinescu 2014]





**Reduction of ID to interleaved MMAP** [Liu, Ihler 2012] + **GDD (Generalized Dual Decomposition)** [Ping,[Liu, Ihler 2015]





❖ **Direct methods for bounding IDs**

**Mini-bucket elimination with valuation algebra (MBE-VA)** [Dechter 2000, Maua 2012]

**Information relaxation by minimum sufficient information set (IR-SIS)** [Nilsson 2001, Yuan 2010]

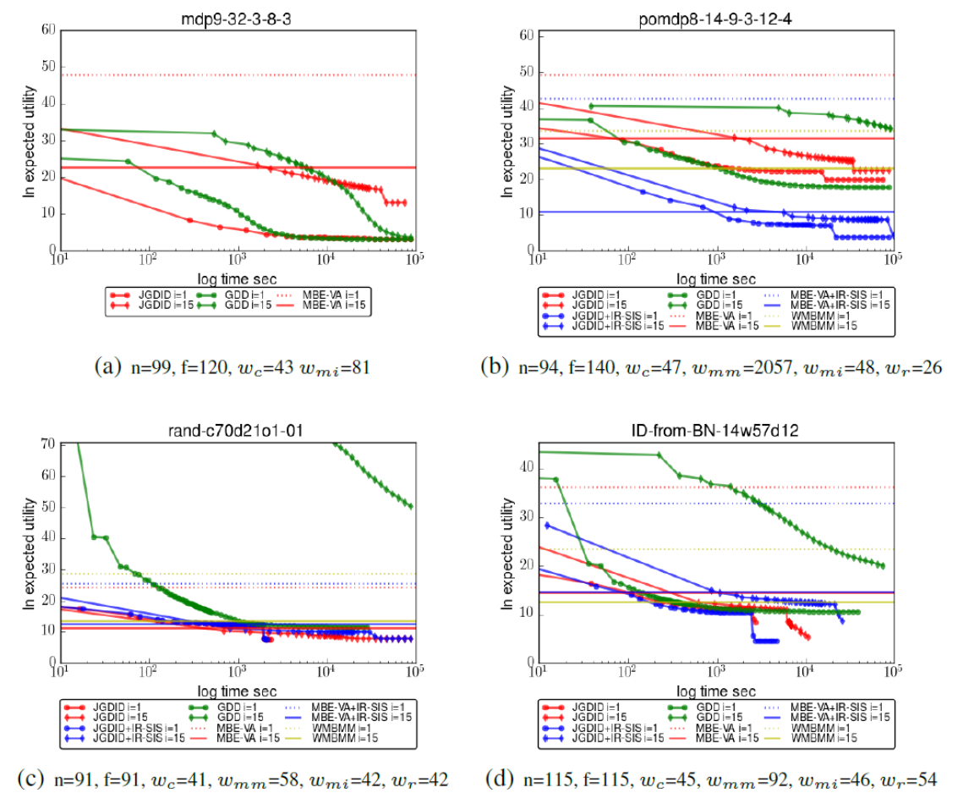
## Experiments

❖ **Benchmarks**

| Domain | n | f | k | s | $w$ |
|---|---|---|---|---|---|
| FH-MDP | 25, 110, 170 | 30, 143, 170 | 2, 3, 5 | 4, 7, 9 | 5, 25, 43 |
| FH-POMDP | 15, 51, 96 | 18, 61, 140 | 2, 2, 3 | 3, 5, 9 | 10, 27, 47 |
| RAND | 22, 57, 91 | 22, 57, 91 | 2, 2, 2 | 3, 3, 3 | 6, 18, 41 |
| BN | 54, 100, 115 | 54, 100, 115 | 2, 2, 2 | 6, 8, 10 | 18, 28, 45 |

Each domain has 10 problems instances
Tables shows min, median, max of the followings
n: number of chance and decision variables,
f: number of probability and utility functions,
k: maximum domain size,
s: maximum scope size,
w: constrained induced width

❖ **Upper bounds**



(a) n=99, f=120, $w_c$=43 $w_{mi}$=81

(b) n=94, f=140, $w_c$=47, $w_{mm}$=2057, $w_{mi}$=48, $w_r$=26

(c) n=91, f=91, $w_c$=41, $w_{mm}$=58, $w_{mi}$=42, $w_r$=42

(d) n=115, f=115, $w_c$=45, $w_{mm}$=92, $w_{mi}$=46, $w_r$=54

❖ **Algorithms**

| | | |
|---|---|---|
| Proposed algorithm | JGDID<br>JGDID+ IR-SIS | Iterative<br>i-bound=1, 15 |
| Translation based methods | WMBMM with MMAP translation | Non-iterative<br>i-bound=1,15 |
| | GDD with interleaved MMAP translation | iterative<br>i-bound=1, 15 |
| Direct methods | MBE-VA<br>MBE-VA+IR-SIS | Non-iterative<br>i-bound=1, 15 |

❖ **Average Quality**

- average of [[best UB/ UB of algorithm] ( 0<= quality <= 1)

| Algorithm | FH-MDP | FH-POMDP | RAND | BN |
|---|---|---|---|---|
| JGDID+IR-SIS(i=1) | NA | 0.88 | 0.87 | 0.99 |
| JGDID+IR-SIS(i=15) | NA | 0.76 | 0.85 | 0.64 |
| JGDID(i=1) | 0.88 | 0.38 | 0.86 | 0.89 |
| JGDID(i=15) | 0.49 | 0.38 | 0.85 | 0.64 |
| MBE-VA+IR-SIS(i=1) | NA | 0.03 | 0.01 | 0.00 |
| MBE-VA+IR-SIS(i=15) | NA | 0.54 | 0.46 | 0.15 |
| MBE-VA(i=1) | 0.00 | 0.00 | 0.00 | 0.00 |
| MBE-VA(i=15) | 0.40 | 0.29 | 0.46 | 0.17 |
| GDD(i=1) | 0.87 | 0.03 | 0.11 | 0.24 |
| GDD(i=15) | 0.22 | 0.11 | 0.15 | 0.05 |
| WMBMM(i=1) | 0.00 | 0.00 | 0.01 | 0.01 |
| WMBMM(i=15) | 0.01 | 0.23 | 0.35 | 0.24 |